

27/01/2021

La transferència antropomòrfica de la Intel·ligència Artificial



Pot un robot tenir emocions o entendre les dels humans? En aquest article Jordi Vallverdú, professor del Departament de Filosofia ens explica amb matisos una narrativa molt propícia del cinema de ciència-ficció, però que cada vegada va adquirint més forma d'una "llegenda urbana" que s'enriqueix de la necessitat humana de percebre la realitat en clau emocional i d'arribar a creure que les màquines senten, el cel s'enfada i que la nostra tortuga ens entén.

"Intel·ligència Artificial emocional", "Robots emocionals", o "computació afectiva" són termes que cada cop escoltem més sovint en els diversos mitjans de comunicació. Tots remeten a una mateixa realitat: la recerca que uneix les emocions i les màquines, projectant un escenari que a molta gent preocupa, el de les màquines que entenguin completament les emocions humanes o, encara més, tinguin emocions per sí mateixes. Ara bé, què hi ha del cert en tot això? Realment hi ha màquines que tinguin emocions? Malgrat el que haguem pogut veure, la resposta és clara i directa: no. Tot i així, nombrosos estudis es publiciten tot ensenyant sistemes d'IA o robòtica que pretesament experimenten emocions. Aquesta apropiació indeguda de capacitats, és alhora el resultat d'una transferència antropomòrfica.

D'altra banda, assistim a un segon gran problema en el disseny de màquines que captin emocions humanes o, pretesament, les reproduïxin: partir d'un model esbiaixat o directament erroni del que és en conjunt de processos afectius que envolten la cultura humana. Per començar aquesta camps de recerca estan dominats completament pels homes occidentals que projecten la seva visió pròpia i personal dels entorns afectius, de manera que es poden identificar molts errors de disseny en relació a assumptes de gènere.

Aquesta manca de treballs interdisciplinaris provoca l'ús de models simplistes sobre la realitat de les emocions, com ara l'ús del model d'Ekman de 6 emocions bàsiques, el qual ni és correcte, ni universalment distribuït...però s'ajusta correctament a les expectatives de programació i de treball d'un enginyer. A banda d'això, tal i com recull el nostre article, també es produeix una sobrevaloració dels beneficis de la implementació de màquines emocionals, quelcom no del tot clar, i menys en l'estat actual de (im)maduresa dels sistemes d'IA. I, paradoxalment, la situació oposada també es produeix amb similars efectes nocius: la menysvaloració del poder que les màquines amb habilitats emocionals tenen per la societat humana. Sabem que els dissenys de veus artificials amables però assertives aconseguirien que els usuaris segueixin les seves indicacions, més que no pas que en el cas de veus amb signes d'inseguretat.

En darrer lloc, també és fàcilment identificable un altre problema: l'atribució incorrecta d'emocions a màquines que realitzen operacions banals. Pensem en un robot que arrufa les celles i tanca la boca...podem dir que està enfadat? O més aviat que ens ho sembla? Atesa la intrínseca necessitat dels humans de percebre la realitat en clau emocional, necessitem creure que les màquines senten, que el cel s'enfada quan plou o que la nostra tortuga és una mascota que ens entén.

A mode de conclusió cal admetre, doncs, que en el disseny de màquines i programari que té característiques emocionals es produeixen nombrosos errors, biaixos i assumpcions infundades. Tenint en compte que les emocions ens fan ser justament que si som, com a individus i com a espècie, és de vital importància controlar de quina manera les introduïm en les màquines que ens acompanyen. I ara, si us hem despertat l'interès, llegireu l'article. Pura emoció.

Jordi Vallverdú

Àrea de Lògica i de Filosofia de la Ciència.

Departament de Filosofia.

Universitat Autònoma de Barcelona.

jordi.vallverdu@uab.cat

Referències

Jordi Vallverdú, Valentina Franzoni and Alfredo Milani. 2019. **Errors, Biases and Overconfidence in Artificial Emotional Modeling**. In *Proceedings of WI '19: IEEE/WIC/ACM International Conference on WebIntelligence (WI '19 Companion)*, New York, NY, USA, 86-90 pages. <https://doi.org/10.1145/3358695.3361749>

[View low-bandwidth version](#)